

# DOCUMENT RESUME

ED 235 203

TM 830 616

AUTHOR Houser, Ronald L.; And Others  
 TITLE An Alternate Procedure to Obtain Ability Estimates in Latent Trait Models.  
 PUB DATE Apr 83  
 NOTE 8p.; Paper presented at the Annual Meeting of the American Educational Research Association (67th, Montreal, Quebec, April 11-15, 1983).  
 PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)  
 EDRS PRICE MF01/PC01 Plus Postage.  
 DESCRIPTORS Academic Achievement; Educational Testing; Item Banks; \*Latent Trait Theory; \*Measurement Techniques; \*Models; Research Methodology; \*Statistical Analysis; \*Test Reliability  
 IDENTIFIERS \*Fixed Parameter Approach; \*Linkage Analysis; Rasch Model

## ABSTRACT

This report describes a procedure that promises to improve the stability, accuracy, and efficiency of the employment of latent trait models and an application of the procedure to the Rasch model. Data were collected from the Portland Public Schools Level Tests administered to 25,740 students. Since each of the 173 items (chosen from the total number of items tested after consideration of the approximately 50% overlap of items from one level to the next) could be calibrated twice by both the linking procedure and the fixed parameter procedure, the same data set was used to compare the relative efficiency of calibration within each procedure and then to compare the efficiency between procedures. This allowed the focus of each metric to be somewhat different but still provided a test of the calibrating efficiency. The differences in the distributions between each pair of item calibrations can be viewed as the amount of error in each calibrating procedure; no statistical or practical difference between these two distributions were found. A final distribution was computed by subtracting the linking difference from the fixed parameter difference for each item. This distribution, representing the relative superiority of one calibration procedure over the other showed no significant difference between the two procedures. Thus, efficiency of calibrating items through a linking approach or through the fixed parameter procedure is essentially the same. (PN)

\*\*\*\*\*  
 \* Reproductions supplied by EDRS are the best that can be made \*  
 \* from the original document. \*  
 \*\*\*\*\*

ED235203

AN ALTERNATE PROCEDURE TO  
OBTAIN ABILITY ESTIMATES  
IN LATENT TRAIT MODELS

A Paper Presented at the  
AERA Convention  
in  
Montreal, Canada  
April 1983

by  
Ronald L. Houser  
Walter E. Hathaway  
George S. Ingebo

Portland (OR) Public Schools

U.S. DEPARTMENT OF EDUCATION  
NATIONAL INSTITUTE OF EDUCATION  
EDUCATIONAL RESOURCES INFORMATION  
CENTER (ERIC)

☒ This document has been reproduced as  
received from the person or organization  
originating it.

☐ Minor changes have been made to improve  
reproduction quality.

• Points of view or opinions stated in this docu-  
ment do not necessarily represent official NIE  
position or policy.

"PERMISSION TO REPRODUCE THIS  
MATERIAL HAS BEEN GRANTED BY

R. L. Houser, W. E.

Hathaway, & G. S. Ingebo

TO THE EDUCATIONAL RESOURCES  
INFORMATION CENTER (ERIC)."

## An Alternate Procedure to Obtain Ability Estimates in Latent Trait Models

### Introduction

In recent years latent trait measurement models have been employed to meet a variety of measurement requirements. School systems, state departments of education, colleges and universities, and test publishing companies have developed procedures and support systems to enable them to benefit from the "population independent/equal interval" measurement that is offered by such latent trait models as the Rasch model. These institutions have all faced the difficult tasks of linking and scaling to a common metric. The following describes a procedure that promises to improve the stability, accuracy and efficiency of the employment of such models and an application of the procedure to the Rasch model.

The Rasch model essentially has two variables (achievement of the student and difficulty of the items) which must be estimated simultaneously. In the conventional approach to Rasch calibrating an item, as developed by Ben Wright and his colleagues, the counts of the number of students with each different raw score are combined with the counts of the number of students answering each item correctly to produce initial estimates of the calibrations and the achievement or ability levels. The achievement estimates are then used to "improve" the calibration estimates by requiring conformity to the item counts. Conversely, the achievement estimates are "refined" using the calibration estimates. This procedure is repeated until the average change in the estimates is less than some small value (such as .001).

### "Fixed Parameter" Procedure

The authors propose that a direct substitution of an empirical value for the student achievement parameter can be made for the estimated value when:

- 1) there is a previously developed item bank to which one wishes to scale more items;
- 2) a reasonably reliable test or test system has been developed employing items from the item bank;
- 3) item responses to the new set of items and responses to the test incorporating bank items are obtained concurrently.

When these conditions are fulfilled, it is possible to substitute an existing achievement in the solution of the Rasch model without employing an estimation process for obtaining a value for the student achievement parameter. Thus, this information can be used to transform the counts of students answering each item correctly into an item calibration in a single step.

### Strength of the "Fixed Parameter" Procedure

This procedure provides many advantages over an estimation process. First, item quality can be evaluated regardless of the number or quality of the other items on the field test. Second, there is an independence of the calibration estimates from the achievement estimates since there is no simultaneous solution employed. Third, direct scaling to the item bank metric is accomplished since the score groupings are already in the desired metric. Fourth, the efficiency of calibrating items is increased since only 20% as many resources are needed as are needed in a complex linking procedure. Fifth, scaling is accomplished on an individual item basis rather than by blocks of items as employed in linking processes. Finally, because of the reasons just cited, there is less estimation in the system, and thus a decrease in the standard error of calibration.

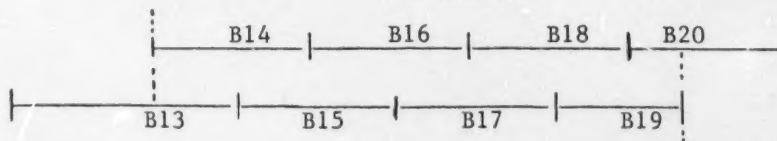
### Purpose

The purpose of this paper is to demonstrate that item calibrations obtained from this "fixed parameter" procedure are at least as consistent as those obtained from the more "conventional" linking procedure.

### Method

Portland Public Schools has developed a series of "Level Tests" based on an item bank which has been carefully Rasch calibrated over the past six years. The researchers chose the mathematics test series B from Fall 1981 and Spring 1982 for our data set in this study. The tests are designed so that they can be administered on a functional basis, therefore, they increase in difficulty from level B 13 to B 20 and overlap by 50% of the items with the next higher level as indicated in Figure 1.

Figure 1



There were a total of 25,740 students who were administered these tests and they were distributed across the levels as indicated in Table 1.

Table 1

| <u>Test Level</u> | <u>Students</u> | <u>Items</u> | <u>RIT Range</u> |
|-------------------|-----------------|--------------|------------------|
| B13               | 851             | 55           | 164 - 194        |
| B14               | 1754            | 55           | 173 - 203        |
| B15               | 3227            | 55           | 189 - 218        |
| B16               | 4895            | 55           | 196 - 225        |
| B17               | 6128            | 55           | 203 - 234        |
| B18               | 5308            | 50           | 217 - 245        |
| B19               | 3183            | 45           | 231 - 255        |
| B20               | <u>394</u>      | 45           | 229 - 277        |
|                   | 25740           |              |                  |

With the approximately 50% overlap of items from one test level to the next, the number of items that could be calibrated twice was reduced to 173.

Portland uses a metric named RIT (Rasch Un I T) with a center of 200 and a logit equal to 10 RIT units. The calibration ranges (given in RIT units) of each of the B series level tests is also presented in Table 1.

The ability estimate needed in the fixed parameter procedure was obtained from the achievement score the students received from this test. This scale score was based upon the item calibrations developed over several years with a variety of measuring instruments and which therefore was not dependent upon the item responses in this particular administration.

In the linking procedure, each test was independently calibrated and then the linking constants were developed between each successive pair of tests. The linking constants and the number of items involved in each link are presented in Table 2.

Table 2

| <u>Link</u> |           | <u>Constant</u> | <u>#Linking Items</u> | <u>Average Calibration</u> |             |
|-------------|-----------|-----------------|-----------------------|----------------------------|-------------|
| <u>From</u> | <u>To</u> |                 |                       |                            |             |
| B13         | B14       | - 10.9          | 25                    | B13 = 205.5                | B14 = 194.6 |
| B14         | B15       | - 10.1          | 25                    | B14 = 206.1                | B15 = 196.0 |
| B15         | B16       | - 9.9           | 23                    | B15 = 203.9                | B16 = 194.0 |
| B16         | B17       | - 11.5          | 28                    | B16 = 205.5                | B17 = 194.1 |
| B17         | B18       | - 11.0          | 22                    | B17 = 206.3                | B18 = 195.3 |
| B18         | B19       | - 9.8           | 18                    | B18 = 206.2                | B19 = 196.4 |
| B19         | B20       | - 9.6           | 19                    | B19 = 204.7                | B20 = 195.0 |

The linking constants were then applied to the entire set of items for each test. Test level B15 was chosen as the center of the distribution because the average of the item bank calibrations for level B15 is 200.4.

In order to answer the research question, the authors chose a design that would control as many extraneous variables as possible. Since each of the 173 items can be calibrated twice by both the linking procedure and the fixed parameter procedure, the same data set was used to compare the relative efficiency of calibration within each procedure and then to compare the efficiency between procedures. This allowed the focus of each metric to be somewhat different but still provided a test of the calibrating efficiency. Table 3 represents the means and standard deviations of the item calibrations for the odd numbered levels vs the ones obtained on the even numbered level tests by calibrating procedure.

Table 3

|                           | <u>Linking Procedure</u>    |           |          | <u>Fixed Parameter Procedure</u> |           |          |
|---------------------------|-----------------------------|-----------|----------|----------------------------------|-----------|----------|
|                           | <u><math>\bar{X}</math></u> | <u>SD</u> | <u>N</u> | <u><math>\bar{X}</math></u>      | <u>SD</u> | <u>N</u> |
| Levels B13, B15, B17, B19 | 215.12                      | 21.35     | 173      | 214.55                           | 20.81     | 173      |
| Levels B14, B16, B18, B20 | 215.06                      | 21.31     | 173      | 214.46                           | 20.67     | 173      |

A "t" test yielded no significant differences between each pair of calibration means at the .01 level of significance.

Another way to view this data is to describe the distributions of the differences between each pair of item calibrations. These means and standard deviations are presented in Table 4.

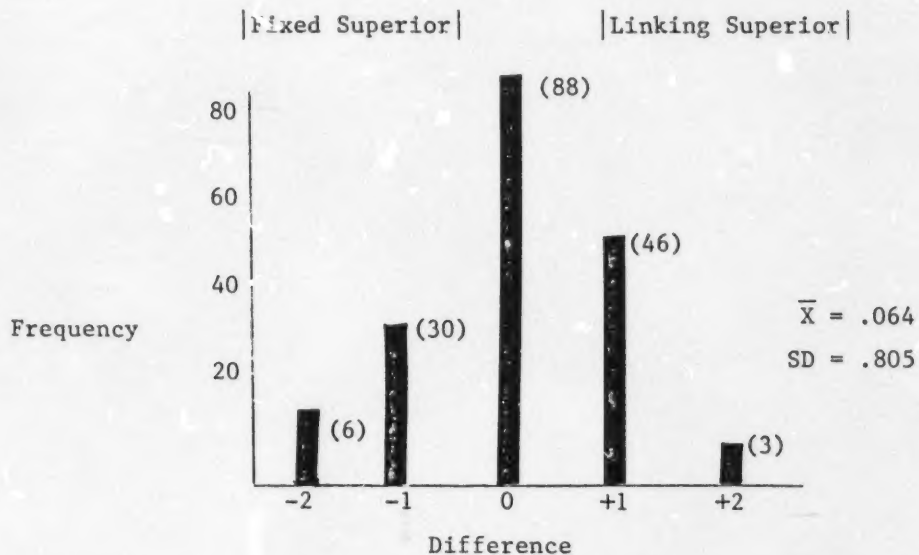
Table 4

|                   | <u><math>\bar{X}</math></u> | <u>SD</u> | <u>N</u> |
|-------------------|-----------------------------|-----------|----------|
| Linking Procedure | 1.93                        | 1.50      | 173      |
| Fixed Procedure   | 1.99                        | 1.50      | 173      |

These differences can be viewed as the amount of error in each calibrating procedure. There is no statistical or practical difference between these two distributions.

A final distribution was computed by subtracting the linking difference (in calibrations) from the fixed parameter difference (in calibrations) for each item. This distribution represents the relative superiority of one calibrating procedure over the other. If the mean is zero, then both procedures are equally effective in calibrating this set of items. The mean, standard deviation and histogram are presented in Figure 2.

Figure 2



Again, there is no significant difference between the two procedures. The units expressed here are in RITs (ie.  $\frac{1}{10}$  logit).

### Conclusions

This research project has demonstrated that the relative efficiency of calibrating items through a linking approach or through the fixed parameter procedure is essentially the same. This has profound practical implications.

Although it is necessary to use more expensive and time-consuming linking methods in initially developing a Rasch-scaled item bank, once a relatively complete, accurately calibrated and scaled bank has been created, the "fixed parameter" method permits the expansion of that bank at a much more rapid pace and lower cost. This expansion may be desired for alternate form development, creating improved items, obtaining better goal coverage, or better coverage of high and low extremes of measurement. The procedure allows a tremendous increase in the efficiency of the number of field tests required to calibrate an item.